Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest eigenvalues: a simulation study.

Pocrnic *et al. Genet Sel Evol* (2019) 51:75
https://doi.org/10.1186/s12711-019-0516-0

GSE Genetics Selection Evolution

RESEARCH ARTICLE Open Access

Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest eigenvalues: a simulation study

Ivan Pocrnic*, Daniela A. L. Lourenco, Yutaka Masuda and Ignacy Misztal

Abstract

Background: The dimensionality of genomic information is limited by the number of independent chromosome segments (M_e), which is a function of the effective population size. This dimensionality can be determined approximately by singular value decomposition of the gene content matrix, by eigenvalue decomposition of the genomic relationship matrix (GRM), or by the number of core animals in the algorithm for proven and young (APY) that maximizes the accuracy of genomic prediction. In the latter, core animals act as proxies to linear combinations of M_e. Field studies indicate that a moderate accuracy of genomic selection is achieved with a small dataset, but that further improvement of the accuracy requires much more data. When only one quarter of the optimal number of core animals are used in the APY algorithm, the accuracy of genomic selection is only slightly below the optimal value. This suggests that genomic selection works on clusters of M_e.

Results: The simulation included datasets with different population sizes and amounts of phenotypic information. Computations were done by genomic best linear unbiased prediction (GBLUP) with selected eigenvalues and corresponding eigenvectors of the GRM set to zero. About four eigenvalues in the GRM explained 10% of the genomic variation, and less than 2% of the total eigenvalues explained 50% of the genomic variation. With limited phenotypic information, the accuracy of GBLUP was close to the peak where most of the smallest eigenvalues were set to zero. With a large amount of phenotypic information, accuracy increased as smaller eigenvalues were added.

Conclusions: A small amount of phenotypic data is sufficient to estimate only the effects of the largest eigenvalues and the associated eigenvectors that contain a large fraction of the genomic information, and a very large amount of data is required to estimate the remaining eigenvalues that account for a limited amount of genomic information. Core animals in the APY algorithm act as proxies of almost the same number of eigenvalues. By using an eigenvalues-based approach, it was possible to explain why the moderate accuracy of genomic selection based on small datasets only increases slowly as more data are added.

Background

Genomic best linear unbiased prediction (GBLUP) is a common tool for genomic analysis in animal and plant breeding [1]. Its basic form is equivalent to single nucleotide polymorphism (SNP) BLUP [2] and assumes an identical distribution of all SNP effects [1, 3, 4]. When not all the individuals are genotyped, a special version of GBLUP called single-step GBLUP (ssGBLUP) can merge pedigree and genomic relationships into a single matrix [5]. The advantage of GBLUP (and especially ssGBLUP) is simplicity, since existing models and BLUP software can be reused just by changing a relationship matrix.

*Correspondence: pocrnic.ivan@gmail.com
Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA

GBLUP and ssGBLUP have become popular methodologies for the genetic evaluation of livestock. Although Bayesian variable selection methods [2, 6] were found to be more accurate with small datasets, their advantage seemed to be lost with large reference populations [7]. Daetwyler et al. [8] showed that selection of SNPs via BayesB outperformed GBLUP only if the number of quantitative trait loci (QTL) was small compared to the number of independent chromosome segments (M_e). Therefore, if the amount of phenotypic data is small, SNPs that are selected by tagging large QTL segments can improve accuracy by reducing the number of parameters to estimate. Karaman et al. [7] found that the advantage of BayesB over GBLUP fades with large datasets. Consequently, when the amount of information is sufficient to estimate most of the segments, selection of SNPs is no longer beneficial. Although selection of SNPs is possible with GBLUP [9, 10], its application is difficult in complex multitrait models, such as those used for commercial genetic evaluations.

There are several formulas to determine M_e. The first formula reported by Stam [11] is based on the number of chromosome junctions in a fixed size population with random mating, i.e. $4N_eL$, where N_e is the effective size of the population and L is the genome length in Morgan. By taking selection into account, Hayes et al. [12] reduced that number to $2N_eL$, and Goddard [4] reduced that number even further to $2N_eL/\log(4N_eL)$. Assuming typical values for N_e (100) and L (30) in Holstein dairy cattle, according to these three formulas, M_e would be equal to 12,000, 6000, and 600, respectively.

Pocrnic et al. [13] related M_e to the dimensionality of the genomic relationship matrix (GRM). For large populations that are genotyped with many SNPs, N_eL, $2N_eL$, and $4N_eL$ corresponded approximately to the number of eigenvalues that explained 90, 95, and 98% of the GRM variation, respectively. To determine which number of eigenvalues maximizes the accuracy of genomic selection, they applied ssGBLUP with a GRM inverted by the algorithm for proven and young (APY) [14], which computes a sparse generalized inverse while indirectly assuming M_e as derived in Misztal [15]. The accuracy of prediction was maximized for a range of N_e when the assumed dimensionality was approximately $4N_eL$. However, the accuracy was only marginally lower when the assumed dimensionality was $2N_eL$ or N_eL. Pocrnic et al. [16] found similar results when analyzing field datasets for dairy and beef cattle, pigs, and chickens and estimated the M_e at ~ 10,000 to 15,000 in cattle and ~ 4000 in pigs and chickens. Although the theory of genomic prediction by chromosome segments is interesting, it seems to be incomplete. Assuming that all chromosome segments are independent and approximately of equal size, Daetwyler et al. [8, 17], Goddard [4], Goddard et al. [18] presented several formulas to estimate accuracy of genomic selection based on heritability, M_e, and the size of the reference population. However, in a meta-analysis using field datasets, their formulas had little predictive power [19].

If all the segments had approximately the same size, assuming half the optimal dimensionality in the APY (the largest eigenvalues that explained 98% of the GRM variation/2) would lead to half the reliability compared with using full dimensionality. However, using half of the optimal number as core animals reduced the reliability by less than 2%, and using only a third of that number reduced the reliability by less than 5% [13, 16]. Therefore, the decrease in reliability was tiny with both simulated and field datasets. In Pocrnic et al. [16], approximately 25% of the eigenvalues explained more than 90% of the genetic variation in the GRM. This suggests that genomic selection by GBLUP (and SNP BLUP) can also be seen as being based on estimates of eigenvalues of GRM. The first purpose of our study was to determine the distribution of eigenvalues in a GRM as well as the GBLUP accuracy when only the top eigenvalues of the GRM are considered. The second purpose was to determine if the optimum number of core animals in the APY algorithm is more related to the number of independent chromosome segments or to the number of top eigenvalues.

Methods

Data simulation

Data for this study were generated using the QMSim software [20]. Each of the simulated scenarios was replicated five times. The initial historical population consisted of 1250 generations with a gradual decrease in size from 5000 to 1000 breeding individuals and then an increase to 25,015 breeding individuals with equal sex ratio, non-overlapping generations, random mating, no selection, and no migration, in order to create a bottleneck and initial linkage disequilibrium (LD) and to establish mutation-drift balance in the population. Then, 10 discrete, recent generations with N_e of ~ 40 were simulated by random mating of 1000 females and 10 males per generation, which resulted in 6000 genotyped individuals in generations 8 to 10. Phenotypes for individuals from generations 8 and 9 were simulated with an overall mean as the only fixed effect and with assumed heritabilities of 0.1, 0.3, 0.6, and 0.9. Scenarios with a heritability of 0.6 were replicated by simulating half (3000) and twice (12,000) the number of genotyped animals. To keep N_e consistent across scenarios with increasing or decreasing numbers of animals, the number of breeding males per generation was fixed at 10. The simulated genome was assumed to have 10 chromosomes of equal length of 100 cM each; 3000 biallelic and randomly distributed

QTL affected the trait, with allelic effects sampled from a gamma distribution as predefined in the QMSim software. The recurrent mutation rate of the markers and QTL was assumed to be 2.5×10^{-5} per locus per generation [21]. The first generation of the historic population had 50,000 evenly allocated biallelic SNPs with equal allele frequencies.

Model and GRM matrices

GBLUP was used for the analysis with the following model $\mathbf{y} = \mathbf{1}\mu + \mathbf{u} + \mathbf{e}$ with $\text{var}(\mathbf{u}) = \mathbf{G}\sigma^2_{\mathbf{u}}$ and $\text{var}(\mathbf{e}) = \mathbf{I}\sigma^2_{\mathbf{e}}$, where $\mathbf{y}$ is a vector of phenotypes, μ is a simple mean, $\mathbf{u}$ is a vector of animal effects, $\mathbf{e}$ is a vector of residuals, $\mathbf{G}$ is a GRM, $\sigma^2_{\mathbf{u}}$ is the additive variance set to result in the desired heritability, and $\sigma^2_{\mathbf{e}}$ is the residual variance.

GBLUP was run with three options for the GRM. For the first option, a standard GRM was constructed as in VanRaden [1]:

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{2\sum p_j(1-p_j)},$$

where $\mathbf{Z}$ is a matrix of allele content centered for allele frequency and p_j is the allele frequency for marker j. For the second option, a reduced-rank GRM was constructed based on $\mathbf{G} = \mathbf{UDU}'$, where $\mathbf{U}$ is a matrix of eigenvectors and $\mathbf{D}$ is a diagonal matrix of eigenvalues arranged from the highest to the lowest value. Then, a GRM restricted to r eigenvalues and eigenvectors ($\mathbf{G}_{\text{eig}}$) was constructed as $\mathbf{G}_{\text{eig}} = \mathbf{UD}_r\mathbf{U}'$, where $\mathbf{D}_r$ includes only the r largest eigenvalues in $\mathbf{D}$. To enable inversion in GBLUP, 0.01 $\mathbf{I}$ was added to both $\mathbf{G}$ and $\mathbf{G}_{\text{eig}}$ for full rank. This method is equivalent to using the largest singular values in the SNP-BLUP design matrix ($\mathbf{Z}$). As a third option, the inverse of the GRM was derived using APY ($\mathbf{G}^{-1}_{\text{APY}}$) as in Misztal [15]:

$$\mathbf{G}^{-1}_{\text{APY}} = \begin{bmatrix} \mathbf{G}^{-1}_{cc} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{G}^{-1}_{cc}\mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}^{-1}_{nn} \begin{bmatrix} -\mathbf{G}_{nc}\mathbf{G}^{-1}_{cc}\mathbf{I} \end{bmatrix},$$

where c and n designate core and noncore animals, respectively, in blocks of $\mathbf{G}$ and

$$\mathbf{M}_{nn} = \text{diag}\{m_{nn,i}\} = \text{diag}\left\{g_{ii} - \mathbf{g}_{ic}\mathbf{G}^{-1}_{cc}\mathbf{g}_{ci}\right\}.$$

The inverse is sparse and requires only the dense inverse of the block of GRM for core animals.

Computations

Standard GRM were calculated for the three populations (3000, 6000, and 12,000 genotyped animals) and replicated five times. Then, the number of eigenvalues that explained approximately 10, 30, 50, 70, 90, 95, and 98% of the variance in the GRM was computed; the fraction was defined as $\text{tr}(\mathbf{Dr})/\text{tr}(\mathbf{D})$. Subsequent computations were performed only on the 6000-animal population. GBLUP was run using standard GRM ($\mathbf{G}$), $\mathbf{G}_{\text{eig}}$, and $\mathbf{G}^{-1}_{\text{APY}}$. For $\mathbf{G}^{-1}_{\text{APY}}$, the same number of eigenvalues as for $\mathbf{G}_{\text{eig}}$ was used as number of core animals. Core animals were chosen randomly from all available genotypes.

Validation

Two methods for assessing accuracy were applied. The first method calculated a realized accuracy as the correlation between the genomic estimated breeding value and the simulated breeding value for animals from the last generation without phenotypes. The second method was based on prediction error variance (PEV) that was calculated in a training set of animals. Validation was done on exactly the same animals as in the first method, but this time those animals were completely excluded from the GBLUP equations. The number of validation animals varied per scenario and was 1000, 2000, or 4000.

The accuracy for animal i (acc_i) based on PEV is calculated as follows:

$$\text{acc}_i = \sqrt{1 - \frac{\text{PEV}_i}{\sigma^2_a g_{ii}}} = \sqrt{1 - \frac{\text{LHS}^{ii}}{\sigma^2_a g_{ii}}},$$

where LHS^{ii} is the diagonal term of the inverse of the left-hand side of the mixed-model equations corresponding to animal i. The same accuracy can be represented as:

$$\text{acc}_i \approx \sqrt{1 - \frac{\alpha}{\alpha + d^p_i + d^g_i}} \approx \sqrt{1 - \frac{\alpha}{\alpha + 1 + d^g_i}},$$

where $\alpha = \sigma^2_e/\sigma^2_a$ is the ratio of residual to animal genetic (a) variance and d^p_i and d^g_i are the effective number of records per individual for phenotypic and genomic information, respectively [22–24]; with one phenotype per animal, $d^p_i \approx 1$. If the amount of genomic information is calculated for animals with phenotypes only, the approximate accuracy for young animals from the same population but with no phenotypic information will be:

$$\sqrt{1 - \frac{\alpha}{\alpha + \overline{d^g_i}}},$$

where $\overline{d^g_i}$ is the average amount of genomic information based on a d^g_i of a training population and is common for all the validation animals. The d^g_i of a training population was based on PEV that are calculated by a direct inversion of the corresponding left-hand side of the mixed-model equation for training animals using the BLUPF90 software [25].

These two methods can be compared because they both result in a measure of accuracy based on the whole population rather than on individuals.

Results and discussion

Figure 1 shows the eigenvalue profiles for 3000, 6000, and 12,000 genotyped animals. The number of eigenvalues that explained 30, 50, 70, 90, 95 and 98% of the total genomic variation ranged from 15 to 16, 45 to 49, 113 to 130, 357 to 453, 585 to 804, and 964 to 1495, respectively. Standard deviations across replicates were negligible. When varying the number of genotyped animals, the number of eigenvalues that explained a given percentage of the variance did not change much for lower percentages of explained variance, and the change was more marked for higher percentages. For lower percentages of explained variance (10 to 50%), the number of eigenvalues was relatively small (3 to 50). For higher percentages, the number of eigenvalues was more variable. For example, the number of eigenvalues that explained 90% of the GRM variance ranged from about 900 for a population of 3000 genotyped animals to 1800 for 12,000 animals. Based on Stam [11], Pocrnic et al. [13] reported that approximately $4N_eL$ eigenvalues explained 98% of the variance, but their study assumed a population much larger than $4N_eL$, and the eigenvalue profile undergoes compression at higher percentages for smaller populations. The logarithm of the number of eigenvalues explaining 30 to 90% of the GRM variance increased almost linearly.

The accuracy of GBLUP with the standard **G** increased with increased heritability as expected and was used as a benchmark for the $\mathbf{G}_{\text{eig}}$ and $\mathbf{G}^{-1}_{\text{APY}}$ methods. Average accuracy (± standard error) values were 0.69 ± 0.03, 0.79 ± 0.01, 0.90 ± 0.01, and 0.96 ± 0.00 for heritabilities of 0.1, 0.3, 0.6, and 0.9, respectively.

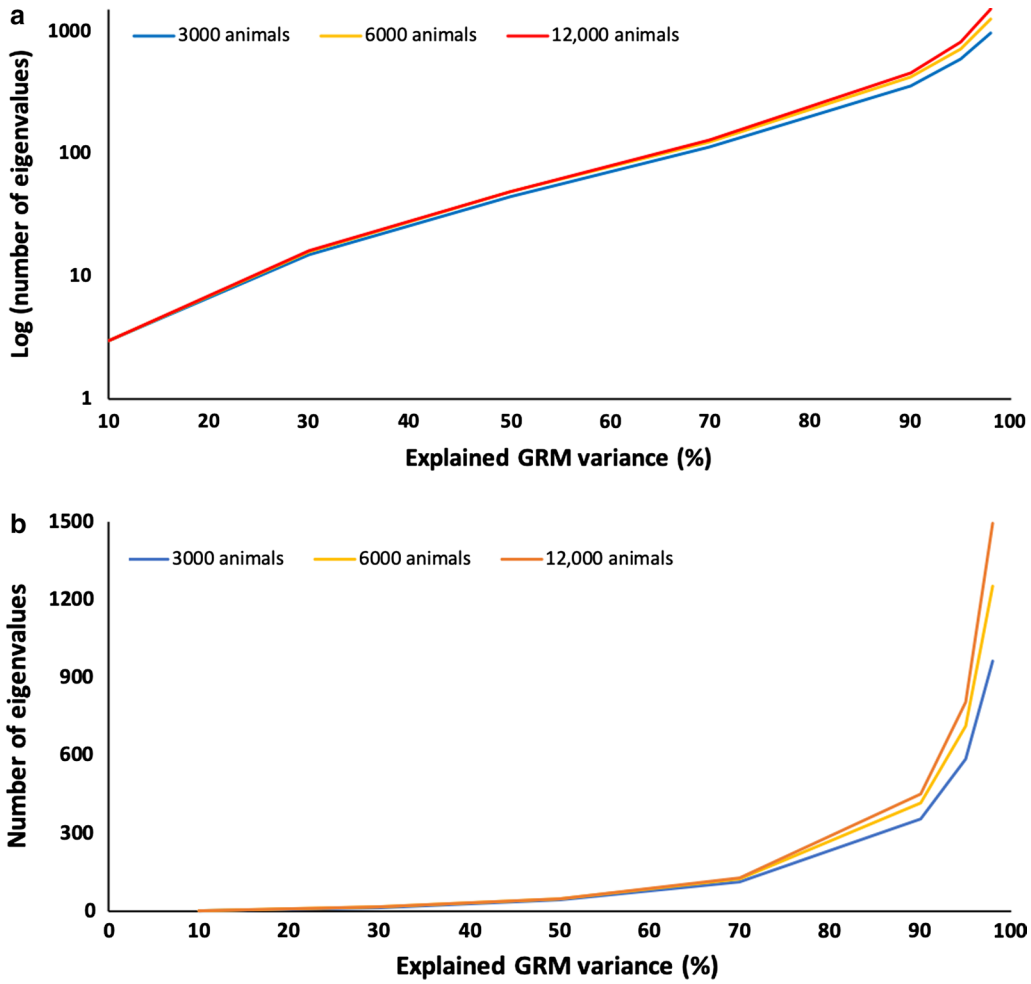

Fig. 1 Eigenvalue profiles for explained variance of the genomic relationship matrix (GRM). Eigenvalues are expressed as either the log of the number (**a**) or the number itself (**b**) for simulated populations of 3000, 6000, and 12,000 genotyped animals

For a heritability of 0.6 and half the number of animals (3000), average accuracy was reduced to 0.87 ± 0.01; with twice the number of animals (12,000) it increased to 0.92 ± 0.01.

The accuracy of GBLUP with $\mathbf{G}_{eig}$ relative to the percentage of explained GRM variance is shown in Fig. 2 and the corresponding number of eigenvalues in Fig. 3 for heritabilities of 0.1, 0.3, and 0.9 for 6000 genotyped animals. For a heritability of 0.1, accuracy stops increasing at ~ 70% of the explained variance and for a heritability of 0.3, it stops increasing at ~ 90% of the explained variance. For a heritability of 0.9, it continues to improve up to 98% of the explained variance. For all heritabilities, accuracy at 98% of the explained GRM variance was the same as for GBLUP with a standard **G**. Figure 4 shows the eigenvalues on a logarithmic scale for 6000 genotyped animals and heritabilities of 0.1, 0.3, and 0.9 and includes points beyond which eigenvalues are smaller

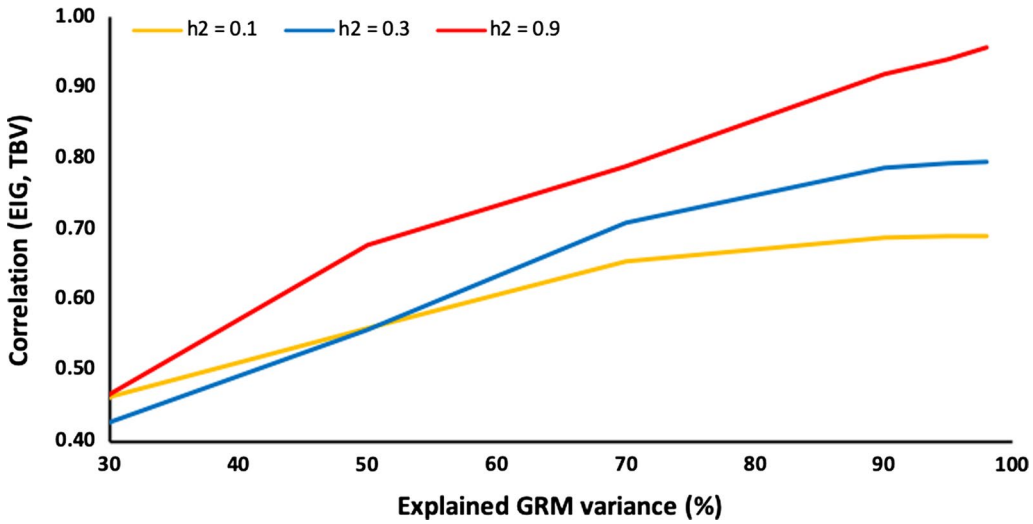

Fig. 2 Accuracy of the genomic relationship matrix (GRM) restricted by eigenvalues based on the percentage of explained GRM variance (EIG) and heritability (h^2). Accuracy is measured as the correlation between genomic estimated breeding values obtained with EIG and simulated breeding values (TBV). Heritability (h^2) was 0.1, 0.3, or 0.9 for a population of 6000 genotyped animals

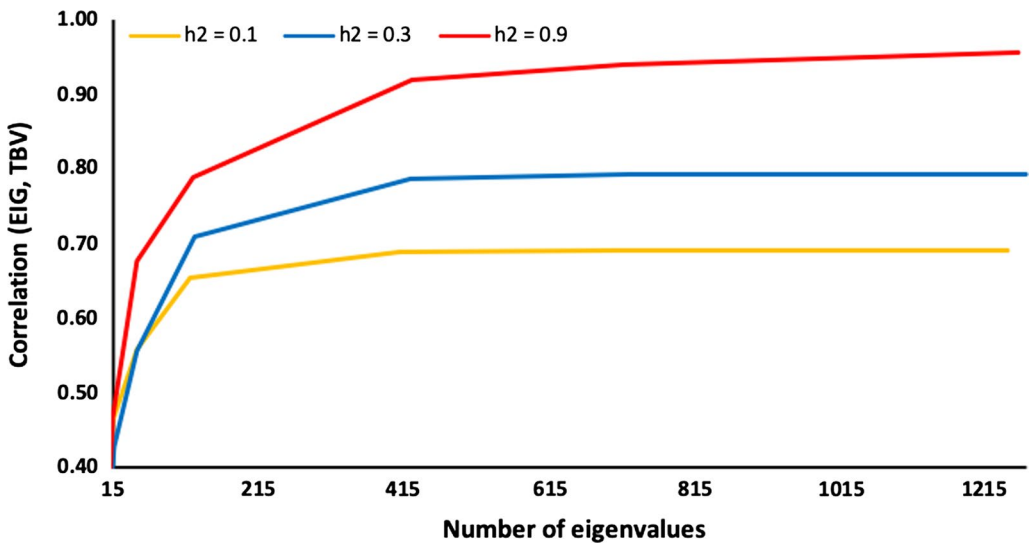

Fig. 3 Accuracy of the genomic relationship matrix restricted by eigenvalues (EIG) based on number of eigenvalues and heritability (h^2). Accuracy is measured as the correlation between genomic estimated breeding values obtained with EIG and simulated breeding values (TBV). Heritability (h^2) was 0.1, 0.3, or 0.9 for a population of 6000 genotyped animals

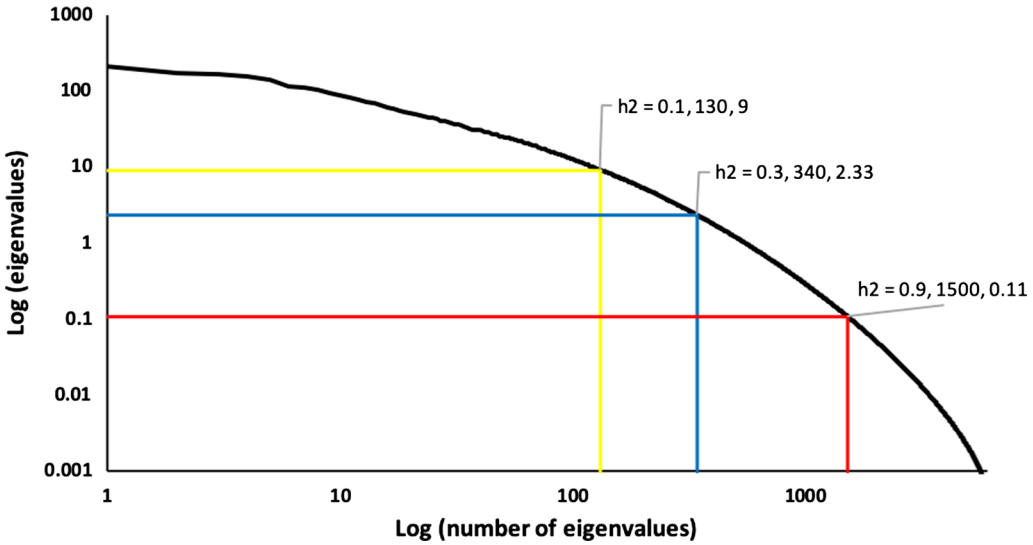

Fig. 4 Relationship between logs of eigenvalues and numbers of eigenvalues for a population of 6000 genotyped animals. Specific curve points beyond which the eigenvalues are smaller than the ratio of residual to animal genetic variance are noted for heritabilities (h^2) of 0.1, 0.3, and 0.9. The values shown after h^2 are the number of eigenvalues at specific curve points and the variance ratios at given h^2

than the variance ratio α; details on the computation are provided in the Appendix. These eigenvalues are likely to affect accuracy, whereas smaller eigenvalues are likely to be ignored. For a heritability of 0.1, the point is approximately a log(eigenvalue) of 130, which corresponds to 70% of the explained GRM variance; the corresponding point is ~ 340 (< 90% of explained variance) for a heritability of 0.3 and ~ 1500 (98–99% of the explained variance) for a heritability of 0.9. These points correspond approximately to the points where the accuracy plateau is reached for $\mathbf{G}_{\text{eig}}$ (Figs. 2 and 3). The lower the heritability (or the smaller the effective information), the fewer eigenvalues are considered, and subsequently the information included in the smaller eigenvalues is ignored. With a higher heritability, the information contained in smaller eigenvalues is included.

The accuracy of GBLUP with $\mathbf{G}_{\text{eig}}$ relative to the number of eigenvalues is shown in Fig. 5 for population sizes of 3000, 6000, and 12,000 and a heritability of 0.6. For the largest population, accuracy is slightly lower at smaller numbers of eigenvalues and slightly higher for larger numbers of eigenvalues. In general, accuracy is expected to be higher with a larger population when a complete relationship matrix is used. However, the largest eigenvalues could correspond to the largest clusters of haplotypes, and those clusters can account for slightly more variation with smaller populations. Accuracy increases when genetically similar animals are part of the reference population; therefore, prediction accuracy for a large population with many animals for which both genotypes and phenotypes are available will improve by including additional information (e.g., herd mates) in the reference population [26]. For all population sizes, differences in accuracy were small. When the amount of phenotypic information is sufficient to estimate the effects due to most of the eigenvalues, accuracy is high and improves little with additional data.

Figure 6 shows the average accuracy of GBLUP with heritabilities of 0.3 and 0.9 for $\mathbf{G}_{\text{eig}}$ and $\mathbf{G}^{-1}_{\text{APY}}$ using the same number of eigenvalues and core animals, respectively, for a population of 6000 genotyped animals. Accuracy is lower for $\mathbf{G}^{-1}_{\text{APY}}$ than for $\mathbf{G}_{\text{eig}}$ at the number of eigenvalues corresponding to 70% of the explained variance but very similar at larger numbers. Using n eigenvalues is almost equivalent to assuming recursion with n animals. Therefore, animal effects for any n animals include almost the same information as the n largest eigenvalues. Sampling variance among the five replicates was larger with $\mathbf{G}^{-1}_{\text{APY}}$ than with $\mathbf{G}_{\text{eig}}$, especially at smaller numbers. The choice of the core animals in the APY algorithm is critical when their number is small but not when it is large [13].

Validation methods used to assess accuracy of GBLUP are compared in Fig. 7. For all heritability levels, accuracy was slightly lower for the method based on average number of effective records than for realized accuracy. The difference was largest for a heritability of 0.3 and smallest for a heritability of 0.9. The method based on average number of effective records can be a useful and simple

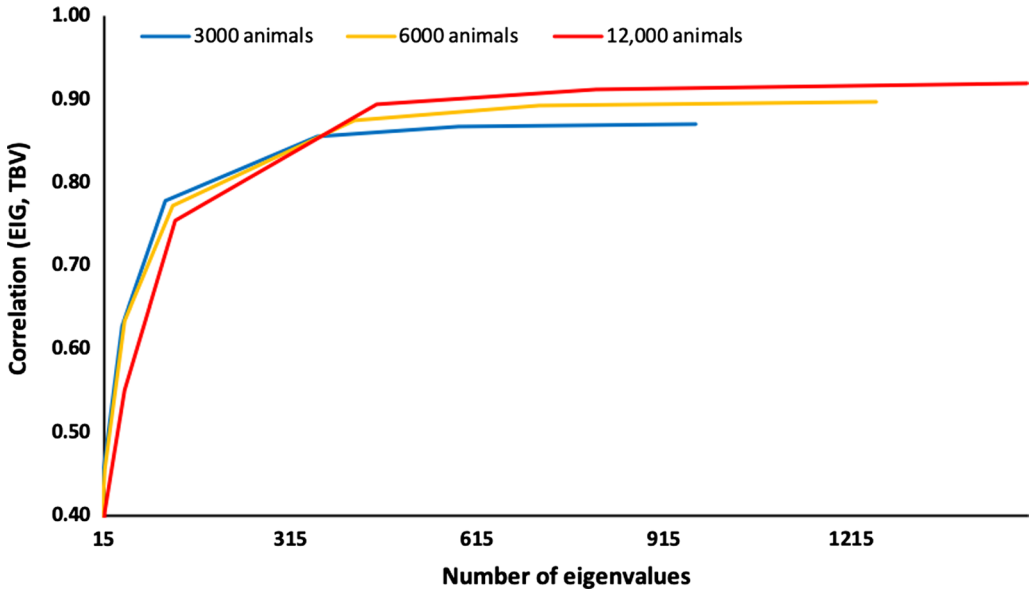

Fig. 5 Accuracy of the genomic relationship matrix restricted by eigenvalues (EIG) based on number of eigenvalues and population size. Accuracy is measured as the correlation between genomic estimated breeding values obtained with the EIG and simulated breeding values (TBV). Population size was 3000, 6000, or 12,000 genotyped animals with a heritability of 0.6

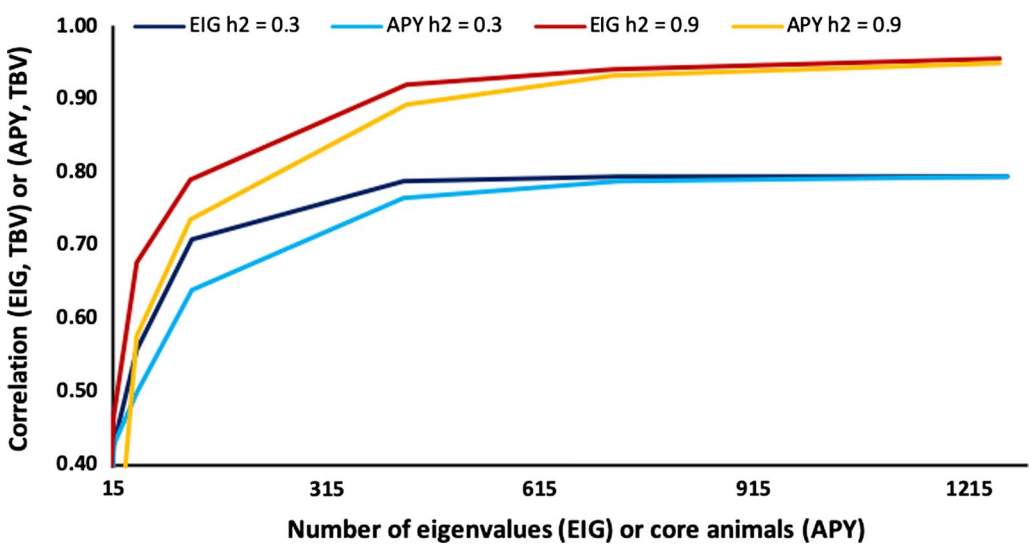

Fig. 6 Accuracy of the genomic relationship matrix either restricted by eigenvalues (EIG) or with the inverse derived by using the algorithm for proven and young (APY) based on number of core animals [15]. Accuracy is measured as the correlation of simulated breeding values (TBV) with genomic estimated breeding values obtained with either EIG or APY. Heritability (h^2) was either 0.3 or 0.9 for a population of 6000 genotyped animals

approximation for population accuracies of validation animals.

In animal breeding programs, approximations of individual accuracy are of interest, but they cannot be derived by inversion because of the large amount of data. Although several approximations exist, those formulas are unclear when evaluations include genomic information [24, 27, 28]. One possibility is to use eigenvalue decomposition of **G** (possible derivations are presented in the Appendix). PEV from the direct inversion of the left-hand side of the mixed-model equation were compared with PEV from the eigenvalue decomposition of **G**

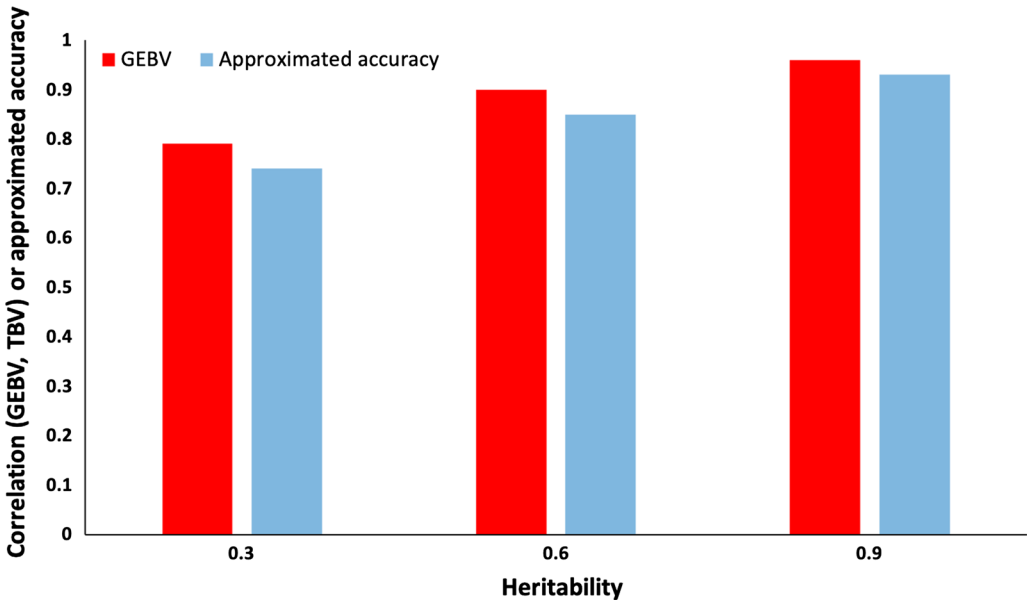

Fig. 7 Comparison of the correlation between genomic estimated breeding values (GEBV) and simulated breeding values (TBV) with accuracy approximated from the average number of effective records. Heritability was 0.3, 0.6, or 0.9, and the simulated population included 6000 genotyped animals

using 2000, 4000, and 8000 genotyped animals that were treated as training animals for validation with heritabilities of 0.1, 0.3, 0.6, and 0.9. For all scenarios, correlations were equal to 1. Meuwissen et al. [29] reported a similar method for obtaining PEV using singular value decomposition for SNP BLUP. Approaches to approximate accuracy are experimental, and further research is needed to evaluate and incorporate these formulas beyond simple GBLUP, especially for ssGBLUP.

It would be useful to derive new formulas on expected genomic accuracies given the heritabilities, the number of genotyped animals and population parameters. According to this study, such an accuracy depends on the fraction of variance explained by subsequent eigenvalues. We attempted to capture that fraction given different effective population sizes and genome lengths. Preliminary studies indicated that the biggest eigenvalues were not affected by genome length, the smallest eigenvalues were affected by population size and all eigenvalues were affected by effective population size. We plan to address this issue in a future study.

Conclusions

The distribution of eigenvalues of the GRM is very uneven, with a small fraction of the largest eigenvalues explaining a large portion of the genetic variation. The accuracy of genomic selection by GBLUP depends on how many eigenvalues can be estimated well, given the amount of information. With a small amount of information, only the effects of the largest eigenvalues are considered, but that small number of eigenvalues can explain a large portion of the genetic variation. Consequently, genomic selection is moderately accurate even with a limited amount of genomic information, and accuracy only increases slowly with larger datasets. Accuracies obtained by GBLUP using the GRM with only n largest eigenvalues and corresponding eigenvectors are similar to using the APY inverse of GRM with recursion on n animals. Subsequently, n animals carry almost the same genomic information as the n largest eigenvalues. Selection by GBLUP is based on clusters of independent chromosome segments and not on individual independent chromosome segments.

Acknowledgements
Editing by Suzanne M. Hubbard is gratefully acknowledged.

Authors' contributions
IP designed the study, analyzed the data and drafted the manuscript; DALL helped with the experimental design, computations, and structure of the manuscript; YM helped with software and computations; IM supervised the project and provided the main ideas. All authors read and approved the final manuscript.

Funding
This research was supported primarily by grants from the American Angus Association, Cobb-Vantress, Genus PIC, Holstein Association USA, Smithfield Premium Genetics, Zoetis, and the U.S. Department of Agriculture's National Institute of Food and Agriculture (Agriculture and Food Research Initiative competitive grant 2015-67015-22936).

Availability of data and materials
The authors state that all data necessary for confirming the conclusions presented in this article are represented fully within the article. In addition, simulation parameter files are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Appendix

Derivation of PEV for individual animals using eigenvalue decomposition.

Let **U** be the eigenvectors and **S** be the eigenvalues of the GRM,

$$\mathbf{G} = \mathbf{USU}'.$$

Assuming full-rank **G**:

$$\mathbf{G}^{-1} = \mathbf{US}^{-1}\mathbf{U}'.$$

The LHS of the mixed-model equation using α as the ratio of residual $\left(\sigma^2_{\mathbf{e}}\right)$ to animal genetic $\left(\sigma^2_{\mathbf{a}}\right)$ variance is:

$$\sigma^2_{\mathbf{e}}\text{LHS} = \mathbf{I} + \alpha\mathbf{G}^{-1} = \mathbf{U}\left(\mathbf{I} + \alpha\mathbf{S}^{-1}\right)\mathbf{U}'$$

with the inverse:

$$\frac{\text{LHS}^{-1}}{\sigma^2_{\mathbf{e}}} = \mathbf{U}\left(\mathbf{I} + \alpha\mathbf{S}^{-1}\right)^{-1}\mathbf{U}' = \mathbf{UFU}',$$

where

$$F = \text{diag}\left\{\frac{1}{1+\frac{\alpha}{s_i}}\right\},$$

and s_i is the eigenvalue i. The effect of eigenvalues on PEV is relative to the variance ratio directly and to the heritability indirectly. Such eigenvalues are especially unimportant if:

$$\frac{\alpha}{s_i} \gg 1 \rightarrow \alpha \gg s_i,$$

or eigenvalues much smaller than α do not matter. For a high heritability, α is smaller and, therefore, smaller eigenvalues matter more. Individual accuracy can be computed as:

$$\frac{\text{LHS}^{ii}}{\sigma^2_e} = u_{i.}\left(\mathbf{I} + \alpha\mathbf{S}^{-1}\right)^{-1}u'_{.i},$$

where $u_{i.}$ is the corresponding eigenvector.

Received: 6 June 2019 Accepted: 4 December 2019
Published online: 12 December 2019

References

1. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.
2. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.
3. Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;177:2389–97.
4. Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica. 2009;136:245–57.
5. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J Dairy Sci. 2010;93:743–52.
6. Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R. Additive genetic variability and the Bayesian alphabet. Genetics. 2009;183:347–63.
7. Karaman E, Cheng H, Firat MZ, Garrick DJ, Fernando RL. An upper bound for accuracy of prediction using GBLUP. PLoS One. 2016;11:e0161054.
8. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. The impact of genetic architecture on genome-wide evaluation methods. Genetics. 2010;185:1021–31.
9. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res (Camb). 2012;94:73–83.
10. Tiezzi F, Maltecca C. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genet Sel Evol. 2015;47:24.
11. Stam P. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genet Res (Camb). 1980;35:131–55.
12. Hayes BJ, Visscher PM, Goddard ME. Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res (Camb). 2009;91:47–60.
13. Pocrnic I, Lourenco DAL, Masuda Y, Legarra A, Misztal I. The dimensionality of genomic information and its effect on genomic prediction. Genetics. 2016;203:573–81.
14. Misztal I, Legarra A, Aguilar I. Using recursion to compute the inverse of the genomic relationship matrix. J Dairy Sci. 2014;97:3943–52.
15. Misztal I. Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics. 2016;202:401–9.
16. Pocrnic I, Lourenco DA, Masuda Y, Misztal I. Dimensionality of genomic information and performance of the algorithm for proven and young for different livestock species. Genet Sel Evol. 2016;48:82.
17. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One. 2008;3:e3395.
18. Goddard ME, Hayes BJ, Meuwissen TH. Using the genomic relationship matrix to predict the accuracy of genomic selection. J Anim Breed Genet. 2011;128:409–21.
19. Brard S, Ricard A. Is the use of formulae a reliable way to predict the accuracy of genomic selection? J Anim Breed Genet. 2015;132:207–17.
20. Sargolzaei M, Schenkel FS. QMSim: a large-scale genome simulator for livestock. Bioinformatics. 2009;25:680–1.
21. Solberg TR, Sonesson AK, Woolliams JA, Meuwissen TH. Genomic selection using different marker types and densities. J Anim Sci. 2008;86:2447–54.
22. Misztal I, Wiggans GR. Approximation of prediction error variance in large-scale animal models. J Dairy Sci. 1988;71:27–32.
23. VanRaden PM, Wiggans GR. Derivation, calculation, and use of national animal model information. J Dairy Sci. 1991;74:2737–46.
24. Misztal I, Tsuruta S, Aguilar I, Legarra A, VanRaden PM, Lawlor TJ. Methods to approximate reliabilities in single-step genomic evaluation. J Dairy Sci. 2013;96:647–54.
25. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). In: Proceedings of the 7th world congress on

genetics applied to livestock production: 19–23 August 2002; Montpellier; 2002.
26. Schultz NE, Weigel KA. An improved genomic prediction model in populations featuring shared environments and familial relatedness. In: Proceedings of the 11th world congress on genetics applied to livestock production: 7–11 February 2018; Auckland; 2018.
27. Liu Z, VanRaden PM, Lidauer MH, Calus MP, Benhajali H, Jorjani H, et al. Approximating genomic reliabilities for national genomic evaluation. Interbull Bulletin. 2017;51:75–85.
28. Edel C, Pimentel ECG, Erbe M, Emmerling R, Götz KU. Short communication: calculating analytical reliabilities for single-step predictions. J Dairy Sci. 2019;102:3259–65.
29. Meuwissen THE, Indahl UG, Ødegård J. Variable selection models for genomic selection using whole-genome sequence data and singular value decomposition. Genet Sel Evol. 2017;49:94.

Publisher's Note